



Identification and mapping of spatial variations in travel choices through combining structural equation modelling and latent class analysis: findings for Great Britain

Kaveh Jahanshahi¹ · Ying Jin¹

Published online: 28 April 2020
© The Author(s) 2020

Abstract

This paper exploits some latest advances in structural equation modelling and latent class analysis for identification and mapping of the spatial variations in travel choices. The approach controls for a wide range of socioeconomic and demographic variables and changes in car fuel prices. The research is focused on employed and self-employed adults, and the method can be readily extended to cover other travellers where such needs arise. The developed methodology enables us to overcome some of the persistent issues that have in the past prevented researchers making a full use of highly correlated and endogenous variables found in good-quality, comprehensive travel surveys at the national or metropolitan scales. Empirical findings from an application of the methodology for Great Britain provide a precise geographical classification of neighbourhoods areas across Britain and reveal the extent to which land use and built form influence commuting travel choices, whilst accounting for residents' self-selection, spatial sorting and endogenous interactions among the explanatory variables. The results are cogent for defining spatially adapted strategies for planning new transport and land use interventions, particularly in areas that are expected to grow the most in the coming decades.

Keywords Structural equation modelling · Latent class analysis · UK National Travel Survey · Commuting travel · Built-up area classification · Travel behaviour · Influence of land use and built form on travel

✉ Kaveh Jahanshahi
kj299@cam.ac.uk

Ying Jin
ying.jin@aha.cam.ac.uk

¹ The Martin Centre for Architecture and Urban Studies, University of Cambridge, 1-5 Scroope Terrace, Cambridge CB2 1PX, UK

Introduction

Overview

This paper aims to identify and map spatial variations in travel choices after controlling for a wide range of socioeconomic and demographic variables and changes in car fuel prices. The research is focused on employed and self-employed adults, and the method can be readily extended to cover other travellers where such needs arise. The methodology exploits latest advances in structural equation modelling (SEM) and latent class Analysis (LCA) of built environment characteristics in order to overcome some of the persistent issues that have in the past prevented researchers making a full use of highly correlated and endogenous variables found in good-quality comprehensive travel surveys.

The paper builds on the growing body of literature on the impact of built environment on travel (e.g. Handy et al. 2005; Van Acker et al. 2007; Mokhtarian and Cao 2008; Gao et al. 2008; Bohte et al. 2009; Cao et al. 2009; Ding et al. 2018) and the recent developments in applying latent clustering techniques to control for heterogeneity among individuals (e.g. Maldonado-Hinarejos et al. 2014; Liao et al. 2015; Delbosc and Naznin 2019; Jahanshahi et al. 2019). The conceptual developments made in a suite of our three recent papers: Jahanshahi et al. (2015) establishes an integrated path diagram for addressing self-selection, spatial sorting, car ownership endogeneity and interactions among trip purposes; Jahanshahi and Jin (2015) incorporates latent class variables into SEM and thus better quantify the effects of different types of built-up areas; Jahanshahi and Jin (2016) embeds random intercept SEM to quantify more precisely the influences of self-selection and spatial sorting in expanded market segmentation and built-up area typology.

The development and testing of the SEM–LCA methodology in this paper has been made possible through the provision by UK Department for Transport of a more comprehensive set of UK National Travel Survey (NTS) data with significantly improved resolution of socioeconomic, demographic, and geographical segmentations through a Transport Research Innovation Grant project funded and supported through an advisory group by UK Department for Transport. The findings are cogent for transport policy making, particularly in the suburban and exurban areas that have good access to the growing metropolitan cores.

Review of literature

There has always been very intense interest in how travel choice vary across different geographies. Some of this interest is placed upon identifying the influences of land use and the built environment on travel choices, since such influences have been considered powerful long term drivers of travel behaviour. Because such influences are mediated through a complex web of interacting demographic, socioeconomic factors concerning the travellers, and through the wider societal context such as cultural values, prices and technology, only very limited progress has been made to date. For those countries that have systematically collected good quality travel survey data, there have been considerable frustration in not being able to make the full range of variables regarding location, personal and household profiles in modelling travel behaviour.

For instance, in the UK the current approach to understanding spatial variations in travel behaviour tends to focus on using deliberately crude categorization such as the population density of the settlements and a NTS-defined land Area Type (for definition of these see

built environment characteristics in Table 1). Headicar (2018) suggest that out of all possible categorisation the NTS Area Type is the preferred way of defining spatial variations of travel behaviour.

Recent years have also seen a growing body of literature that aims to identify more precisely the effects of land use and the built environment on travel demand through better controls for interdependencies between travellers' socioeconomic and demographic profiles, social and cultural attitudes and car ownership. (Handy et al. 2005; Van Acker et al. 2007; Mokhtarian and Cao 2008; Gao et al. 2008; Bohte et al. 2009; Cao et al. 2009; Sun et al. 2009; Robert and Murakami 2010; Silva et al. 2012; Sun et al. 2012; Zegras et al. 2012; Jahanshahi and Jin 2015, 2016; Jahanshahi et al. 2015; Ding et al. 2018). Notably, Gao et al. (2008) analyse the connections between job accessibility, workers per capita, income per capita and cars per capita with census tract data for Sacramento, CA by employing a Structural Equation Model (SEM) to capture endogeneity effects. They find that the error terms of many variables strongly correlate and a multivariate regression model would exaggerate the significance of their influences. Without the application of a SEM, it would not be possible to address such statistical biases that arise from the endogeneity effects.

Our previous papers (Jahanshahi et al. 2015; Jahanshahi and Jin 2015, 2016) have provided a systematic coverage of the SEM applications regarding temporal shifts in the interdependencies, which means that we will focus here on additional literature that are cogent to analyses that combine LCA and SEM.

LCA is a technique to reduce the dimensionality of highly correlated variables such as those exist in the characterisation of complex phenomenon into a tangible list of distinct classes. Recent improvements in statistical methods have made it feasible to characterize heterogeneity among individuals or choices through *latent classes* behind individuals' decisions—e.g. (Maldonado-Hinarejos et al. 2014; Liao et al. 2015; Delbosc and Naznin 2019; Jahanshahi et al. 2019); Notably, de Ona et al. (2013) have used LCA in combination with SEM to classify traffic accidents and identify the main factors affecting accident rates; Beckman and Goulias (2008) have used LCA for classifying immigrants by their travel time, mode choice, and departure time for work to investigate the distinct effects of spatial, social, demographic, and economic influences. More recent example is de Haas et al. (2018) who have used latent class and transition analysis to examine travel pattern classes and transitions between these classes over time.

As complex phenomena go, the built environment in a country is probably one of the best examples for being characterised by highly correlated variables. In the UK NTS survey, the built environment has been described through area type, population density, access to different forms of public transport, etc. (see Table 1). However, there has been few works so far that aim to employ LCA for identifying latent built environment classes. Categorizing geographical locations through latent class analysis can better quantify the built environment influences to inform land use and transport planning and investment. It would therefore appear of both theoretical and policy interest to incorporate LCA in examining the more complex and controversial aspects of the influences of built environment on travel behaviour.

To fill this research gap, our earlier work (Jahanshahi and Jin 2015) has started to develop a SEM with LCA to identify distinct built environment classes in Britain and variations in influences on travel patterns across them. However, this precursor research was restricted by the amount of spatial information that was available at the time. More recently the UK DfT have made more extensive NTS data available in response to this research need. It means that this paper can now make a significant extension to our earlier work by establishing a fully-fledged SEM–LCA model with more distinct (five as opposed to three) latent classes and which, for the first time, will lead to specific mapping of the geographical

variations of the interdependencies. Access to more rigorous data has also enabled us to demonstrate the advantages of exploiting SEM–LCA for spatial classification through comparison with alternatives where a subset of the components (e.g. Area Type) is considered. This is further discussed in Sect. 4.

Methodology

Figure 1 presents a diagram of the combined structural equation modelling—latent class analysis (SEM–LCA) framework. SEM–LCA consists of two sub-models: On the left of the diagram, the LCA characterizes a group of highly correlated indicators depicting the built environment (i.e. area Type, Population Density, and three types of accessibilities to public transport stations—for more detailed explanation, refer to the last two rows of Table 1) as a latent categorical variable. The identification of this latent variable (which we call latent classes for the rest of this paper) is also conditional on socioeconomic characteristics of the travellers. This is then fed into the SEM on the right side of the diagram which is in effect a path model treating socio-economic and fuel price variables as exogenous explanatory variables and car ownership as an intervening variable.

The LCA approach is similar to Factor Analysis (FA): Both LCA and FA are to construct latent variables from observed indicators, but the estimated latent variable is continuous for FA and discrete (or categorical) for LCA. In particular, LCA is focused on grouping the travellers facing distinct patterns of indicators into classes (i.e. identifying distinct latent classes and estimating the probability of latent class membership), and is thus a respondent-centered approach (Wang and Chen 2012).

Model estimation is carried out in two stages: first, we use conditional LCA to identify individuals who reside in different geographical locations by estimating individuals' built environment class membership conditional on their socioeconomic and demographic profile; second, the SEM is used to account for the interdependencies among the residents' socioeconomic

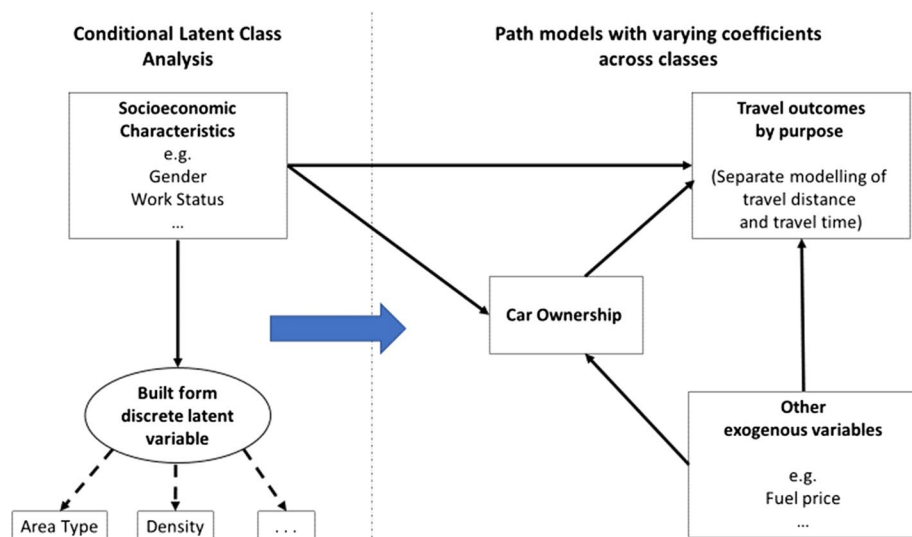


Fig. 1 A SEM–LCA conceptual model for identifying travel choice outcomes

and demographic profiles, their car ownership status, and the different journey purposes in the quantification of the direct and indirect influences on the travel outcomes. The second stage estimation is performed conditional on the class membership which is estimated in the first.

To formulate the first stage, let Y_{ij} be the j th indicator variable (i.e. population density, area type, etc.) of the built form latent class variable, C_i for individual i . As all our indicators are ordered categorical variables, we can formulate the link function by defining an underlying continuous variable, Y_{ij}^* such that

$$Y_{ij} = s | C_i = k \Leftrightarrow \tau_{kj,s} < Y_{ij}^* < \tau_{kj,s+1} \quad (1)$$

where C_i , our latent class variable (i.e. built form), takes values 1, ..., k and $\tau_{kj,s}$ are a set of threshold parameters.

Conditional on regressors X (i.e. socioeconomic characteristics in our case) we can then present the link function as:

$$Y_{ij}^* | C_i = k, x_i = v_{kj} + K_{kj} X_i + \varepsilon_{ij} \quad (2)$$

The normal distribution assumption for ε_{ij} is equivalent to a probit regression for categorical variable Y_{ij} on X_i with the following probability function:

$$\Pr(Y_{ij} = s | c_i = k) = \Phi[(\tau_{kj,s+1} - v_{kj} - K_{kj} X_i)] - \Phi[(\tau_{kj,s} - v_{kj} - K_{kj} X_i)] \quad (3)$$

The class membership probability conditional on X is given by multinomial logistic regression with the following formula:

$$\Pr(C_i = k | X_i) = \frac{\exp(\alpha_k + \gamma_k X_i)}{\sum_{s=1}^k \exp(\alpha_s + \gamma_s X_i)} \quad (4)$$

The joint probability of indicators or observed data likelihood is then given by:

$$\Pr(Y_{i1} \dots Y_{iJ}) = \prod_i \sum_{k=1}^c \Pr(C_i = k) \prod_j \Pr(Y_{ij} = s | c_i = k) \quad (5)$$

EM algorithm is then used for estimating the parameters and class membership where the latent variable C_i is treated as missing data. We first compute the posterior distribution for the latent variable. The posterior conditional joint distribution is calculated as:

$$\Pr(C_i = k | *) = \frac{\Pr(C_i = k) \prod_j \Pr(Y_{ij} = s | c_i = k)}{\sum_{k=1}^c \Pr(C_i = k) \prod_j \Pr(Y_{ij} = s | c_i = k)} \quad (6)$$

which is estimated given the parameters.

Given the class membership, model parameters are then estimated through maximising Eq. (5). The model is solved iteratively until reaching convergence.

Equations (7)–(9) specify the structural equation model which is estimated within each latent class for the second stage of our modelling. The subscript for latent class membership is dropped here for simplicity.

$$Y_{ij} = v_j + K_j X_{ij} + \epsilon_{ij} \quad (7)$$

where Y_{ij} refers to the i th respondent and j th vector of a dependant variable (e.g. travel distance for commuting to work) and X_{ij} is the vector of all individual level covariates. v_j and K_j are the vectors of intercepts and the matrices of regression parameters correspondingly.

ϵ_{ij} is a vector of residuals with a mean of zero and covariance Θ . Where the j th observed dependent variable, Y_{ij} , is a normally distributed continuous variable (i.e. the distance and time travelled by journey purpose), the residual variable ϵ_{ij} is assumed normally distributed. For a dichotomous variable Y_{ij} (i.e. car ownership), a normality assumption for ϵ_{ij} is equivalent to the probit regression for Y_{ij} on X_{ij} .

The observed-data likelihood is given by:

$$\prod_{ij} f_{ij}(Y_{ij}) \quad (8)$$

where f_{ij} is the likelihood function for Y_{ij} .

The expected log likelihood is then maximized with respect to model parameter estimation:

$$\sum_{ij} \log(f_{ij}(Y_{ij})) \quad (9)$$

In order to avoid being trapped in a local maxima for log likelihood estimation, we use many different sets of starting values in the iterative maximization procedure. Mplus, the software we are using for estimating SEM–LCA,¹ by default uses 100 different starting values and then selects the best 10 to run to convergence. The software then reports the maximum likelihood achieved for each converged run. In cases where the maximum likelihood is not replicated, we increase the number of runs to be converged to ensure that the maximized value of the likelihood function is replicated.

Data sources

For this paper we use the National Travel Survey (NTS) data for 2002–2015 which forms a consistent time series of the survey that is reasonably representative for the UK.² We have focussed on employed and self-employed adults between 16 and 64 years old in this paper, because they represent the most mobile group of the population; the approach can be expanded to cover the entire population. There are a total of 1.3 million trips and 11.7 million passenger miles travelled for commuting, shopping and other journeys by the adults. For each journey the NTS provides a household weight to account for non-response and a trip weight for the drop-off in the number of trips recorded by respondents during the course of the survey week, uneven recording of short walks by day of the week and the short-fall in reporting long distance trips. This is to ensure the data is representative of travel of an average week for the population as a whole.

¹ We used explanatory factor analysis, structural equation modelling and mixture modelling commands—for more information on Mplus, refer to <https://www.statmodel.com>; accessed 11 April 2019.

² From 2002 to 2012 the NTS data have maintained excellent stability in data collection methods and variable definition, covering all countries of the UK. However, from January 2013, the coverage of the NTS have been reduced to sample residents of England only. The time series 2002–2015 has been selected in this context as a reasonably representative sample for Great Britain.

The NTS data is organised in nested relational tables of households, individuals, long distance journeys prior to the survey, days within the survey week, journeys made during the survey week, the stages of these journeys and vehicles (Morris et al. 2014). Based on previous NTS analysis and large number of experiments, we have selected all the main attributes for households, individuals and their trip-making (Table 1).

Given its importance to travel choices, we also added fuel prices as a non-NTS variable. The fuel price information is collected from “monthly and annual prices of road fuel and petroleum products” dataset updated monthly by the Department of Business, Energy & Industrial Strategy (<https://www.gov.uk/government/statistical-data-sets/oil-and-petroleum-products-monthly-statistics>, accessed 03/05/2017). The premium unleaded price which is used for our analysis are adjusted by Retail Price Index (RPI) to constant prices for January 2002. The adjusted fuel price data is then linked to the NTS dataset based on the month and year of the survey records for each surveyed household.

Main findings

We summarise the main findings of SEM–LCA as follows:

In Sect. 4.1, we first present the definition of latent built environment classes, with the unconditional and conditional probabilities for individuals to be in each class. Second, we compare the socioeconomic and demographic profiles of residents within the built environment latent classes. Third, within each built environment class, we explore influences on weekly travel distance and travel duration by journey purpose after controlling for interactions among journey purposes as well as endogeneities arising from self-selection, spatial sorting and car ownership.

Section 4.2 compare latent classes with conventional classification by NTS Area Type or population density to demonstrate the advantages of the developed methodology.

Section 4.3 presents the mapping of latent classes and compare that with NTS Area Type in terms of geographical distribution.

Finally in Sect. 4.4, to further validate the model findings, we compare the commuting travel elasticities implied by the SEM–LCA model with those provided in WebTAG.³

Latent classes of the built environment

Out of all the built environment attributes from NTS, five variables—namely “area type”, “population density”, “frequency of local buses”, “walk time to bus stop”, and “walk time to rail station” are found to have large FA loading factors suggesting that they should be included in defining the latent built environment classes. The conditional LCA identifies five latent built environment classes with an entropy of 0.737, which suggests that the latent classes are well defined. Table 2 has listed the five latent classes and our nominated labels for them. The following paragraph describes the classes’ characteristics which have prompted us to the tentative names.⁴

³ WebTAG Unit M2 published by UK Department for Transport provides the guidance on travel demand modelling, including benchmark trip-km elasticity with respect to fuel price.

⁴ The entropy measure is defined on a zero to one scale with the value of one indicating the individuals are perfectly classified into the latent classes and zero not at all. The entropy of over 0.7 is considered as an indication of good classification (Wang and Wang 2012).

Table 1 NTS data: definitions of variables selected for SEM–LCA analysis

Levels of respective variables					
Households (from households table)	Household size: 1 adult >1 adult	Annual income: <£25,000 £25,000–49,999 ≥£50,000	Household head occupation: Manual Skilled manual White collar clerical Professional	Car ownership: No access to car Own or access to one or more than one car	
Employed adults (from individual table)	Gender: Male Female	Work status: Full time (FT) Part time (PT)			
Journeys (from journeys table)	Journey purpose (for outbound purpose): Home-based commuting (HBW) All shopping (Sh) All other purposes (Oth)	Trip frequency: Trips/week	Journey distance: Miles/trip	Journey time: min/trip	
Access to transport services at household location (from household table)	Maximum frequency of local bus services: Level 1: < 1 bus per day Level 2: at least 1 bus/day Level 3: at least 1/hour Level 4: at least 1/1/2 hr Level 5: at least one every quarter hour	Walk to bus stop (min): Level 1: <6 Level 2: 7–10 Level 3: 11–15 Level 4: 16–20 Level 5: 21–30 Level 6: 31–60 Level 7: > 60	Walk time to rail station (min): Level 1: < 6 Level 2: 7–10 Level 3: 11–15 Level 4: 16–20 Level 5: 21–30 Level 6: 31–60 Level 7: > 60		
Built environment characteristics at house- hold location (from post code unit level- PSU-table)	Area type: Central London Inner London Outer London Metropolitan areas outside London Outer Conurbation Urban big > 250,000 population Urban large 100,000–250,000 population Urban medium 25,000–100,000 population Rural areas <25,000 population	Population density (persons/hectare): Level 1: <3.01 Level 2: 3.01–10 Level 3: 10.01–20 Level 4: 20.01–30 Level 5: 30.01–35 Level 6: 35.01–40 Level 7: 40.01–45 Level 8: 45.01–50 Level 9: 50.01–60 Level 10: > 60			

Table 2 Tentative names of the latent classes

Latent class number	Tentative names
LC1	Metropolitan Core Dwellers
LC2	Outer Metropolitan Dwellers
LC3	Suburban Dwellers
LC4	Exurban Dwellers
LC5	Rural Dwellers

Panel 3a of Table 3 shows the unconditional probabilities of individuals belonging to each latent class (LC). Based on the estimated model, classes 1 to 5 contain respectively 8%, 17%, 38%, 22% and 14% of all working adults. Conditional probabilities further reveal the patterns of the latent classes (LCs) benchmarked by the specific characteristics of the built environment (Panel 3b). For example, residents in Latent Class 1 (LC1) consist of not only those from the central, inner and outer London (of respectively 4%, 61%, 21%) but also some from big urban areas (11%)—see Panel 3b-1. The members of this class also reside in the densest areas (see Panel 3b-2) and benefit from the most frequent buses and highest level of accessibility to public transport (see Panel 3b-3 to 3b-5). These characteristics of the residents in this latent class prompt us to label them as ‘Metropolitan Core Dwellers’. Similarly, the dominance of residents from outer London and metropolitan urban in Latent Class 2 (LC2, 59% of this class), the dominance of small urban to big urban and outer conurbation areas in Latent Class 3 (LC3, 92% of this class), the dominance of the small urban and rural areas in Latent Class 4 (LC4, 87% of this class) and the absolute dominance of rural areas in Latent Class 5 (LC5, of 99% of residents) give rise to our labeling them as ‘Outer Metropolitan Dwellers’, ‘Suburban Dwellers’, ‘Exurban Dwellers’ and ‘Rural Dwellers’ respectively.

Figure 2 portrays the composition of LCs by NTS Area Type to give a first sense of geography. Almost all those in central London and 95% of the residents in inner London belong to LC1. However, other area types contain increasingly a mixture of the latent classes, suggesting that it would be necessary to explore in more depth. There is clearly some association between the LCs and the Area Types, but on the other hand considerable differences.

The average population density for each portion of the LC that is within a NTS Area Type is also displayed in Fig. 2 (the white labels). It is clear that the LCs have been able to identify distinct population density levels. For instance, for the Inner London Area Type, the LCs have picked up three parts of distinct population density, respectively at 94.0, 36.6 and 21.6 persons/ha for LC1–3. At the other end of the spectrum, the Rural Area Type has been segmented into LC3–5 respectively for density levels of 44.0, 9.1 and 1.1 persons/ha.

By definition, the LCs also interacts with the socioeconomic and demographic profiles of the residents. In defining the LCs, the conditional LCA estimates socioeconomic and demographic covariates. This in effect controls for self-selection and spatial sorting of the residents, and the model outcomes show that such controls often have a material bearing on the results.

The analysis of demographic and socioeconomic variables are reported through odds ratios with one of the LCs designated as a reference class. This is shown in Table 4 where LC5 is chosen as the reference class. An odds ratio higher than 1.0 indicates a higher likelihood than in the reference class, and vice versa.

Table 3 Unconditional and conditional probabilities for five-class LCA model

Indicators	Latent class				
	1-Metropolitan Core Dwellers (N = 8756)	2-Outer Metropolitan Dwellers (N = 18,573)	3-Suburban Dwellers (N = 41,887)	4-Exurban Dwellers (N = 23,601)	5-Rural Dwellers (N = 15,821)
Panel 3a: unconditional probabilities					
	0.08	0.17	0.38	0.22	0.14
Panel 3b: conditional probabilities					
3b-1: area type					
Central London	0.04	0.00	0.00	0.00	0.000
Inner London	0.61	0.01	0.00	0.00	0.000
Outer London	0.21	0.36	0.00	0.01	0.000
Metropolitan areas outside London	0.03	0.23	0.05	0.01	0.000
Outer conurbation	0.00	0.09	0.19	0.01	0.000
Urban areas > 250,000 population	0.11	0.20	0.24	0.05	0.003
Urban areas 100,000–250,000 population	0.00	0.12	0.21	0.07	0.000
Urban areas 25,000–100,000 population	0.00	0.00	0.28	0.19	0.000
Rural areas < 25,000 population	0.00	0.00	0.01	0.66	0.997
3b-2: population density (person/hectare)					
Under 3	0.04	0.02	0.01	0.28	0.97
3.01–10	0.01	0.01	0.15	0.41	0.03
10.01–20	0.01	0.09	0.26	0.21	0.00
20.01–30	0.01	0.12	0.27	0.08	0.00
30.01–35	0.01	0.08	0.10	0.01	0.00
35.01–40	0.00	0.09	0.09	0.00	0.00
40.01–45	0.01	0.13	0.06	0.00	0.00
45.01–50	0.02	0.14	0.03	0.00	0.00
50.01–60	0.08	0.18	0.01	0.00	0.00
Over 60	0.81	0.13	0.02	0.00	0.00

Table 3 (continued)

Indicators	Latent class				
	1-Metropolitan Core Dwellers (N = 8756)	2-Outer Metropolitan Dwellers (N = 18,573)	3-Suburban Dwellers (N = 41,887)	4-Exurban Dwellers (N = 23,601)	5-Rural Dwellers (N = 15,821)
3b-3: bus frequency					
Less than once a day	0.001	0.000	0.001	0.001	0.049
At least once a day	0.000	0.001	0.007	0.044	0.315
At least once every hour	0.006	0.008	0.116	0.322	0.457
At least once every 30 min	0.073	0.212	0.463	0.482	0.157
At least once every 15 min	0.920	0.779	0.414	0.151	0.022
3b-4: walk time to bus stops					
60 min and more	0.000	0.000	0.000	0.000	0.009
31–60 min	0.000	0.000	0.000	0.001	0.035
21–30 min	0.000	0.001	0.001	0.003	0.035
16–20 min	0.000	0.001	0.001	0.008	0.032
11–15 min	0.005	0.010	0.010	0.022	0.041
7–10 min	0.074	0.107	0.081	0.116	0.116
6 min or less	0.920	0.881	0.906	0.850	0.731
3b-5: walk time to rail station					
60 min and more	0.003	0.028	0.087	0.174	0.507
31–60 min	0.071	0.250	0.287	0.263	0.245
21–30 min	0.147	0.194	0.217	0.165	0.075
16–20 min	0.137	0.141	0.131	0.112	0.051
11–15 min	0.191	0.146	0.117	0.113	0.039
7–10 min	0.274	0.145	0.103	0.105	0.043
6 min or less	0.175	0.095	0.058	0.067	0.041

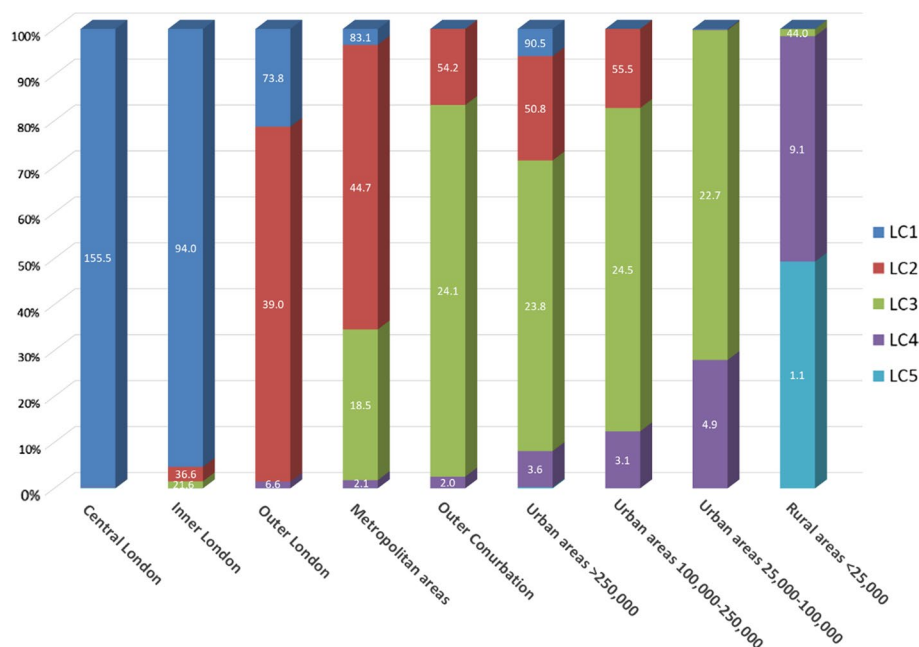


Fig. 2 Composition of built environment latent classes by NTS area type. *Note:* the white labels are the average population density for each area type—LC combination

Unsurprisingly, Table 4 suggest that working adults who reside in LC1 are more likely to come from 1 adult households, and have full time working status; professionals and skilled manual workers are more likely to be found in LC5.

These results reconfirm our earlier SEM work (Jahanshahi et al. 2015). However, the findings from SEM to LCA provide more precise interpretation. For instance, while our earlier work suggests that on average higher income groups tend to live in denser, more urbanized areas, here the SEM–LCA shows precisely the higher likelihood for residing in London and Metropolitan areas (LC1–2). When it comes to comparing LC3–4 with LC5, there is a clearly quantification of the likelihood for high income groups to reside in LC5. These nonlinear patterns cannot be captured from the earlier, path-diagram-based SEM.

Tables 5, 6 and 7 further report the results of the influences upon car ownership, travel distance, and travel time across the LCs. The use of the LCA with SEM provides a unique opportunity to decompose more precisely the influences for each of the demographic and socioeconomic variables by LCs. Furthermore, to identify the additional insights of incorporating a categorical built environment variable in the SEM model, we compare results from SEM to LCA with those from a constrained SEM where the model parameters do not vary across the built environment classes. This constrained SEM is typical of the existing models that do not account for the specific influences of the built environment characteristics.

The model intercepts and coefficients can help to quantify the levels of influences of the demographic and socioeconomic variables in the context of the built environment latent classes. Whilst an intercept represents the average level of car ownership, travel distance or travel time of the reference segment (to aid intuitive interpretation of the model outputs, in all tables we define a reference segment of residents who are female, part time working

Table 4 Odds ratios of demographic and socioeconomic covariates

Covariates	Built environment latent classes				
	1-Metropolitan Core Dwellers	2-Outer Metropolitan Dwellers	3-Suburban Dwellers	4-Exurban Dwellers	5-Rural Dwellers
Male	1.016	0.92***	0.905***	0.918***	Reference latent class
Full time working	1.486***	1.324***	1.252***	1.166***	
1 adult households	2.683***	1.434***	1.281***	1.257***	
Semi- or unskilled manual workers	0.926	0.822***	0.94	0.895*	
Skilled manual workers	0.363***	0.59***	0.766***	0.78***	
Professionals	0.572***	0.582***	0.693***	0.757***	
Low Income HHs	1.021	1.097*	0.992	0.858***	
High Income HHs	1.558***	0.993	0.785***	0.86***	

Base or reference segment is class 2 (medium urban class)

***Significant within 99% CI, **Significant within 95% CI, *Significant within 90% CI

Table 5 Influence of socioeconomic profile on car ownership

Direct effect	Constrained model	LC1	LC2	LC3	LC4	LC5	LC1 versus LC5 Wald test <i>P</i> value
Model threshold for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year		–1.509***	–1.518***	–1.68***	–1.846***	–2.773***	
Male	0.014	–0.015	0.013	0.022	0.043	0.09*	–
Full time working	–0.003	–0.067	–0.015	0.021	–0.015	–0.035	–
1 adult households	–0.523***	–0.355***	–0.5***	–0.509***	–0.543***	–0.735***	0.000
Manual workers	–0.428***	–0.281***	–0.38***	–0.389***	–0.547***	–0.553***	0.018
Skilled manual workers	0.259***	0.332***	0.388***	0.242***	0.048	0.037	0.000
Professionals	0.239***	0.189***	0.244***	0.352***	0.126**	0.261***	0.387
Household income less £25 k	–0.511***	–0.274***	–0.449***	–0.588***	–0.533***	–0.521***	0.000
Household income more than £50 k	0.295***	0.188***	0.454***	0.364***	0.475***	0.317***	0.078
Fuel price	–0.003***	–0.015***	–0.005***	–0.001	0.002	–0.004	0.000

***Significant within 99% CI, **significant within 95% CI, *significant within 90% CI

Table 6 Direct influences on travel distance (in miles) arising from traveller profiles

Direct influence	LC1	LC2	LC3	LC4	LC5	LC1 versus LC5 Wald test <i>P</i> value
Panel 6a. Direct influences on commuting						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	16.4***	12.3***	9.6***	7**	12.7***	
Male	4.9***	10.1***	11.6***	13***	9.4***	0.000
Full time working	11.1***	14.2***	15.9***	19.6***	21.3***	0.000
1 adult households	–1.3	1.6*	2.9***	5***	3.9***	0.000
Semi- or unskilled manual workers	0.8	–1.8**	–3.1***	–3.7***	–6.5***	0.000
Skilled manual workers	–0.7	–1.5*	–4.1***	–8.2***	–8.6***	0.000
Professionals	2**	3.5***	2.8***	2**	1	0.047
Household income less £25 k	–1.5	–2.5***	–4***	–5.8***	–5.4***	0.000
Household income more than £50 k	0.9	3.5***	5.6***	4.7***	4.7***	0.000
Fuel price	–0.1*	–0.1**	–0.1***	0	–0.1**	0.989
Have car in household	1.2	3.2***	5.3***	8.3***	11.1***	0.000
Panel 6b. Direct influences on shopping						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	5.2***	7***	8***	12.4***	20.6***	
Male	–1.2***	–1.8***	–2.6***	–3.5***	–4.7***	0.000
Full time working	–0.6*	–0.4*	–0.4	–1.4***	–1**	0.074
1 adult households	0.6**	1***	0.8***	–0.1	1.1*	0.207
Semi- or unskilled manual workers	–0.1	–0.7**	–1.7***	–1.2***	–0.8	0.511
Skilled manual workers	0.1	–0.6**	–1.5***	–0.2	–1.8***	0.000
Professionals	0.2	0.4	–0.4	1**	–0.9	–
Household income less £25 k	–0.3	–0.8***	–0.6**	–0.5	–1.1**	0.570
Household income more than £50 k	–0.3	–0.3	0.3	0.2	–0.3	–
Fuel price	0	0	0	0	–0.1***	0.000

Table 6 (continued)

Direct influence	LC1	LC2	LC3	LC4	LC5	LC1 versus LC5 Wald test <i>P</i> value
Have car in household	2.2***	3.3***	4.9***	5.3***	6.5***	0.000
Panel 6c. Direct influences on other purposes combined						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	42.4***	27.5***	49.7***	52.2***	50.2***	
Male	4.2**	10.1***	16.5***	18.7***	18.8***	0.000
Full time working	3.6*	–1.8	2.8***	5.4***	6.3***	0.031
1 adult households	18.7***	17.3***	19.1***	21.1***	23.6***	0.000
Semi- or unskilled manual workers	–13.3***	–13.3***	–17***	–21.6***	–30.4***	0.000
Skilled manual workers	–10.3***	–14.7***	–15.8***	–22.1***	–29***	0.000
Professionals	5.8*	14.2***	15.6***	15.8***	14.8***	0.000
Household income less £25 k	–9.2***	–6.8***	–9.8***	–11.8***	–13.8***	0.000
Household income more than £50 k	14.1***	10.7***	15.8***	18***	21.1***	0.000
Fuel price	–0.2	0	–0.3***	–0.3***	–0.1**	0.028

***Significant within 99% CI, **significant within 95% CI, *significant within 90% CI

Table 7 Direct influences on travel time (in min) arising from traveller profiles

Direct influence	LC1	LC2	LC3	LC4	LC5	LC1 versus LC5 Wald test <i>P</i> value
Panel 7a. Direct influences on commuting						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	102.6***	94.5***	76.4***	70.4***	63.3***	
Male	12.2***	14.6***	14.7***	13.7***	8.9***	0.352
Full time working	48.5***	47***	36.1***	40.4***	40.4***	0.000
1 adult households	–17.5***	–8.7***	0	2.3	2.8	0.000
Semi- or unskilled manual workers	11**	–4.3*	–2.2*	–1.2	–10.5***	0.000
Skilled manual workers	–4.5	–8.8***	–9.6***	–13.4***	–15.1***	0.000
Professionals	0.4	2.1	1.2	0.9	–1.3	–
Household income less £25 k	–4.1	–3.2	–4.8***	–6.4***	–6.2***	0.000
Household income more than £50 k	–5.7*	6.8***	6.9***	5.7***	4.4***	0.571
Fuel price	–0.2	–0.3***	–0.2***	–0.2***	–0.2***	0.989
Have car in household	–18.1***	–24.6***	–20.8***	–16.5***	–3.3	0.000
Panel 7b. Direct influences on shopping						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	54.9***	51.4***	56.2***	58.1***	71.5***	
Male	–8.9***	–9.6***	–10***	–10.5***	–11.8***	0.286
Full time working	–3.2**	–3.1***	–3.7***	–5.2***	–4.3***	0.202
1 adult households	2.9**	3.3***	1.5**	0.2	3.1**	0.970
Semi- or unskilled manual workers	–3.1*	–2.5**	–3***	–2.7**	–2.8**	0.575
Skilled manual workers	–3	–2.7***	–3.1***	–2.4***	–4.9***	0.002
Professionals	–1.7	1.2	–2.5***	0.7	–1	–
Household income less £25 k	1.4	–1.2	0.3	–0.8	–0.1	–
Household income more than £50 k	–1.8	–1.5*	0.5	–1.2	–0.1	–
Fuel price	–0.2**	–0.1*	–0.1***	–0.1**	–0.2***	0.989

Table 7 (continued)

Direct influence	LC1	LC2	LC3	LC4	LC5	LC1 versus LC5 Wald test <i>P</i> value
Have car in household	3.2***	−0.5	−0.8	−0.6	0.5***	0.000
Panel 7c. Direct influences on other purposes combined						
Model intercept for the reference segment, which is represented by a female, part time working white collar clerical worker with more than one adults and a household income of 25–50 k per year	226.5***	176.7***	179.2***	178.3***	167.4***	
Male	2.3	14.2***	18.7***	20.8***	19.1***	0.000
Full time working	−13.3***	−13.1***	−10.4***	−2.8	−1	0.000
1 adult households	33.1***	34.3***	34.2***	38.4***	36.8***	0.543
Semi- or unskilled manual workers	−37.2***	−30.5***	−35.3***	−37***	−52.7***	0.157
Skilled manual workers	−25.1***	−35.6***	−35.8***	−42.7***	−48.6***	0.000
Professionals	8.9*	21.7***	21***	19.9***	21.8***	0.000
Household income less £25 k	−10.1**	−12.8***	−15.5***	−20.6***	−19.9***	0.000
Household income more than £50 k	19.9***	15.4***	22.3***	25.8***	30.3***	0.000
Fuel price	−0.5**	−0.3**	−0.6***	−0.6***	−0.4***	0.368
Have car in household	24.3***	29.2***	44.1***	48.9***	59.4***	0.000

***Significant within 99% CI, **significant within 95% CI, *significant within 90% CI

in white collar clerical occupations with more than one adults and a household income of 25–50 k per year), the coefficients indicate how much influence a change in the demographic and socioeconomic profiles has. The rest of the model results provide opportunities to compare the car ownership, travel distances or travel times both within each column (i.e. holding the built environment class constant and decompose the influences of demographic, socioeconomic and car ownership characteristics) and across the columns for each row (i.e. to identify the influence of the built environment given a particular demographic, socioeconomic and car ownership profile). Furthermore, we use Wald tests to examine the statistical significance of differences in influences across built environment latent classes. For simplicity, we report the Wald test results of comparing the first and the fifth classes as the two extremes; this should be sufficient to show if there exists significant variations in influences.

The values for the demographic, socioeconomic and car ownership variable rows are additive within each column for travel distance and travel time. This allows the readers to work out the specific distances or time travelled for an arbitrary type of resident. For car ownership which is estimated by probit regression, the interpretation is not as straightforward. The increase in probability attributed to one-unit increase in a given predictor is also dependent on the reference value of that predictor. This is because the link function for the probit model follows a nonlinear distribution function of the standard normal. Table 5 shows the influences on car ownership across built environment latent classes. In particular, 1 adult households, manual workers, and lowest income groups are more likely to have no car when compared with the reference segment, whilst skilled manual, professionals and high income groups are more likely to have a car in their household. Also, the increase in fuel price is associated with foregoing cars.

The main benefit of SEM-LCA is the insights we get from a comparison across built environment classes which demonstrate significant variations for some socioeconomic groups. Single adult households are more likely to have no access to car but the gap in the level of car ownership is bigger in more rural areas. There is also evidence of significant variations across built environment classes for low income households and manual and skilled manual workers. The difference in car ownership between manual workers and white collar clericals (reference segment) is bigger for exurban and rural dwellers than that in other classes, specifically ‘Metropolitan Core’ areas; for skilled manuals, however, this difference is bigger in ‘Metropolitan Core’ and ‘Outer Metropolitan’ areas while it is not significant in ‘Exurban’ and ‘Rural’ areas. This suggest that this is the white collar clericals who are more prepared to forgo their cars by living in more dense urbanized areas. The difference in car ownership between low, high and medium income households is also significantly larger outside ‘Metropolitan Core’ areas again due to the willingness of the higher income households to have a lower level of car ownership when living in central and inner London and some Metropolitan areas. The model results show that the influence of fuel prices on car ownership varies significantly across classes. Fuel prices seem to have significant influence on car ownership only in “Metropolitan Core” and “Outer Metropolitan” areas and the “Metropolitan Core” (i.e. mainly Central and Inner London) coefficient is three times of that of “Outer Metropolitan” areas. This suggests that in transport models and for evaluating policies, the segmentation by built environment classes is crucial.

Table 6 shows the influence on distance travelled for different purposes across the latent built environment classes. The first line of the model outputs in Panel 6a shows that the reference group differ in their average weekly commuting distances among the five built environment classes: ‘Metropolitan Core’ dwellers travel 16.4 miles per week, Rural dwellers 12.7 miles, and in other areas somewhere between these two extremes. Similarly the first

lines under Panel 6b and 6c in Table 6 show that for shopping and other travel purposes, the more rural the area, the longer the distances travelled which is intuitive. As expected, the reference segment residents commute well below the working adult average of 28.7 miles per week for all classes, but for shopping (for which the average weekly distances travelled is 10.4) they travel shorter than the average in more urbanized areas and longer in the rest.

The general patterns of smaller coefficients for the “Metropolitan Core Dwellers” and “Outer Metropolitan Dwellers” (i.e. relative to its model intercept), and the large ones for other latent classes indicates that the variations in travel distance and duration across socio-economic groups is smaller and that the built environment explains much of the travel in the first two classes; for other areas, however, demographic and socioeconomic profiles have much bigger role to play. For instance, the commuting distance coefficients for high income households (Households with income more than £50 k) in the ‘Metropolitan Core’ and ‘Outer Metropolitan’ classes are respectively 0.9 (and not significant) and 3.5, which shows that by virtue of the higher income, such commuters travel more relative to the reference segments (i.e. by 5.4% and 28% respectively). By contrast, commuters from high income households in ‘Exurban’ and ‘Rural’ areas travel respectively 67% (coefficient 4.7 divided by intercept 7) and 37% (4.7/12.7) more. This pattern is mirrored by the commuting distances for commuters from households with less than 25 k income per year. Similarly, households with no cars in ‘Metropolitan Core’ areas travel similar distance to those with car, whilst those in ‘Rural’ areas travel 11.1 miles less. The rest of built environment classes sit in order between these two extremes.

The results are intuitively correct and they provide a substantially more robust set of quantifications of the influences upon distance travelled by working adults. For instance, many existing models simply suggest that those households with no cars tend to travel much shorter distances than those with cars. However, when we take account of the variations in travel distance across built environment classes, then we see that this is mainly the case in ‘Exurban’ and ‘Rural’ areas but not in ‘Metropolitan Core’ areas.

Table 7 shows the influence on travel time for different purposes across the latent built environment classes. Unlike travel distance which is longer for those from reference group residing in the two extreme classes (i.e. ‘Metropolitan Core’ and ‘Rural’ dwellers), the intercept values in Table 7 suggest that, with the exception of shopping trips, residing in less dense areas is generally associated with longer travel in time. For instance, those live in the ‘Metropolitan Core’ and ‘Outer Metropolitan’ areas spend 102.6 and 94.5 min per week respectively whilst those in ‘Rural’ areas spend 63.3 min for commuting. These results show that the influence is not monotonic; the commuting distance is shorter for ‘Suburban’ and longer for ‘Metropolitan Core’ and ‘Rural’ areas. In terms of commuting time, the longest travel times occur in “Metropolitan Core and Outer Metropolitan” areas, with the rest of area types having broadly similar travel times.

Table 7 suggests that the built environment influence is larger for commuting trips when compared with shopping and other travel purposes (with the exception of Travel time for other purposes which significantly longer for ‘Metropolitan Core’ dwellers). Among the most significant influences, part-time workers spend more time for commuting in ‘Other Metropolitan’ than ‘Rural’ areas. This is also the case for manual and skilled manual workers. The most striking difference is for car ownership with those with no access to car tend

Table 8 Goodness of fit statistics: constrained model versus model allowing variations across classes

	AIC
Travel distance model	
Constrained model	1,506,408
Model allowing variations across classes	1,505,235
Travel time model	
Constrained model	1,975,133
Model allowing variations across classes	1,974,520

to travel relatively 18.1 min, 24.6 min, and 20.8 min longer when residing in ‘Metropolitan Core’, ‘Outer Metropolitan’, and ‘Suburban’ areas respectively.

At the first glance these results might seem unexpected. However, it can be better understood through comparing against influences on travel distance reported in Table 6. While a typical main stream full time workers tend to live closer to their workplace in denser urban areas, skilled and unskilled manual workers and those with no access to car tend to make shorter commuting distances in more rural areas. This can be explained by the latter groups’ reliance on public transport specifically when they reside in denser urban areas. We also suspect the recent surge in house prices may have played a role in this pattern, as the low and medium income households have increasingly found it hard to secure housing in dense, job rich areas.

The goodness of fit statistics prove the better fit when the model allows variations in influences across latent classes (see Table 8). This confirms our previous findings of the significant influence of built environment upon travel. The goodness of fit of the model when benchmarked against conventional area type classification is reported in Sect. 4.2.

A comparison of travel choice identification: latent class versus area type and population density

Table 9 compares the distribution of the standard travel outcomes across our estimated Latent Classes or LCs (panel a) with those across conventional NTS Area Types (panel b) and population density bands (panel c). The comparison clearly demonstrate that far more distinct identification of travel choices can be achieved through the latent class classifications. In addition, the latent classes have identified more distinct behavioural responses when compared with travel outcome variations across area types and population density bands. This is in spite of the fact that travel outcomes themselves do not contribute in defining the LCs—as explained above, conditional latent classes are defined based on built form indicators and conditional on socioeconomic characteristics.

For instance, Table 9 (panel a) shows that, for classification by latent classes, the average travel distance is ranging from 6.9 miles to 11.6 miles with a uni-direction progression in the distribution. This range is smaller for population density (7.1 miles to 10.9 miles) and not uniform for area type classifications—e.g. Outer London is associated with the travel distance of 9.2 miles, the second longest after Rural areas.

Table 9 Comparison of travel outcome distribution across different spatial classification**(a) Latent classes**

	Average travel distance (miles)	Average travel time (minutes)	% of private mode	% of public Transport
LC1	6.9	32.7	35%	59%
LC2	8.0	28.1	64%	33%
LC3	8.3	23.5	80%	18%
LC4	9.9	23.5	84%	14%
LC5	11.6	24.1	88%	11%

(b) NTS area types

	Average travel distance (miles)	Average travel time (minutes)	% of private mode	% of public Transport
Central London	4.6	28.3	12%	78%
Inner London	6.2	33.8	26%	67%
Outer London	9.2	35.2	48%	50%
Metropolitan areas outs	6.7	24.5	75%	21%
Outer Conurbation	7.5	23.8	83%	14%
Urban areas >250k popu	8.5	24.7	76%	21%
Urban areas 100k-250k p	8.3	22.7	79%	18%
Urban areas 25k-100k po	9.0	22.8	79%	19%
Rural areas <25k populat	10.8	23.8	87%	12%

(c) Population density bands

	Average travel distance (miles)	Average travel time (minutes)	% of private mode	% of public Transport
>50	7.1	29.9	43%	38%
35-50	8.0	26.0	66%	19%
20-35	8.3	24.2	71%	14%
5-20	9.0	23.7	76%	10%
<5	10.9	24.1	83%	6%

Table 10 Goodness of fit statistics: constrained model versus model allowing variations across classes

	AIC
Travel distance model	
Model (a)—area type	809,024
Model (b)—latent class	702,845
Travel time model	
Model (a)—area type	1,274,337
Model (b)—latent class	1,168,718

Further, to illustrate the advantages of adopting the latent class methodology, derived in this paper, for modelling travel outcomes, we compared two sets of models: model (a) where the observed NTS Area Types are used as classes and the influences on travel are allowed to vary across those and model (b) where instead of area types, we used the derived latent classes through allocating each individual to their most likely latent class.⁵ Table 10 shows the comparison results for travel distance and travel time models which shows significantly better explanatory power for model (b).

Spatial mapping of the latent classes

How do the LCs spread in geographic space? The SEM–LCA results provide, for the first time, a new opportunity to map the distinct patterns of travel behavior using consistent NTS time series data. Conditional LCA provides the class membership likelihood of each sampled individual in the NTS sample. Here the UK National Transport Model (NTM) zones (which cover all Great Britain) has made it possible to map the LCs for the first time. Figure 3 (panel b) reports the results with eight types of areas: five types to represent areas that are dominated by one latent built environment class (with likelihood of over 70% of residents matching that particular built environment class) and three in-between types acknowledging the existence of highly mixed composition of two adjacent class types (where neither type has a probability of more than 70%)—this is particularly the case outside LC1. In addition, Fig. 3 (panel a) presents the spatial distribution of area types as a reference for comparison. The comparison clearly demonstrates more distinctive patterns offered by the latent classes. For instance, NTS classification of area types has defined Newcastle as a metropolitan area with its immediate surrounding area as outer conurbation. Latent class analysis, however, has offered more detailed classification of three types (i.e. LC2, LC3, and LC4) with better distinction and clear spatial connections.

⁵ The alternative is to weight individuals by their class membership probability which would have given better goodness of fit. However, we use the simpler most likely class membership to allow like with like comparison to area type classifications.

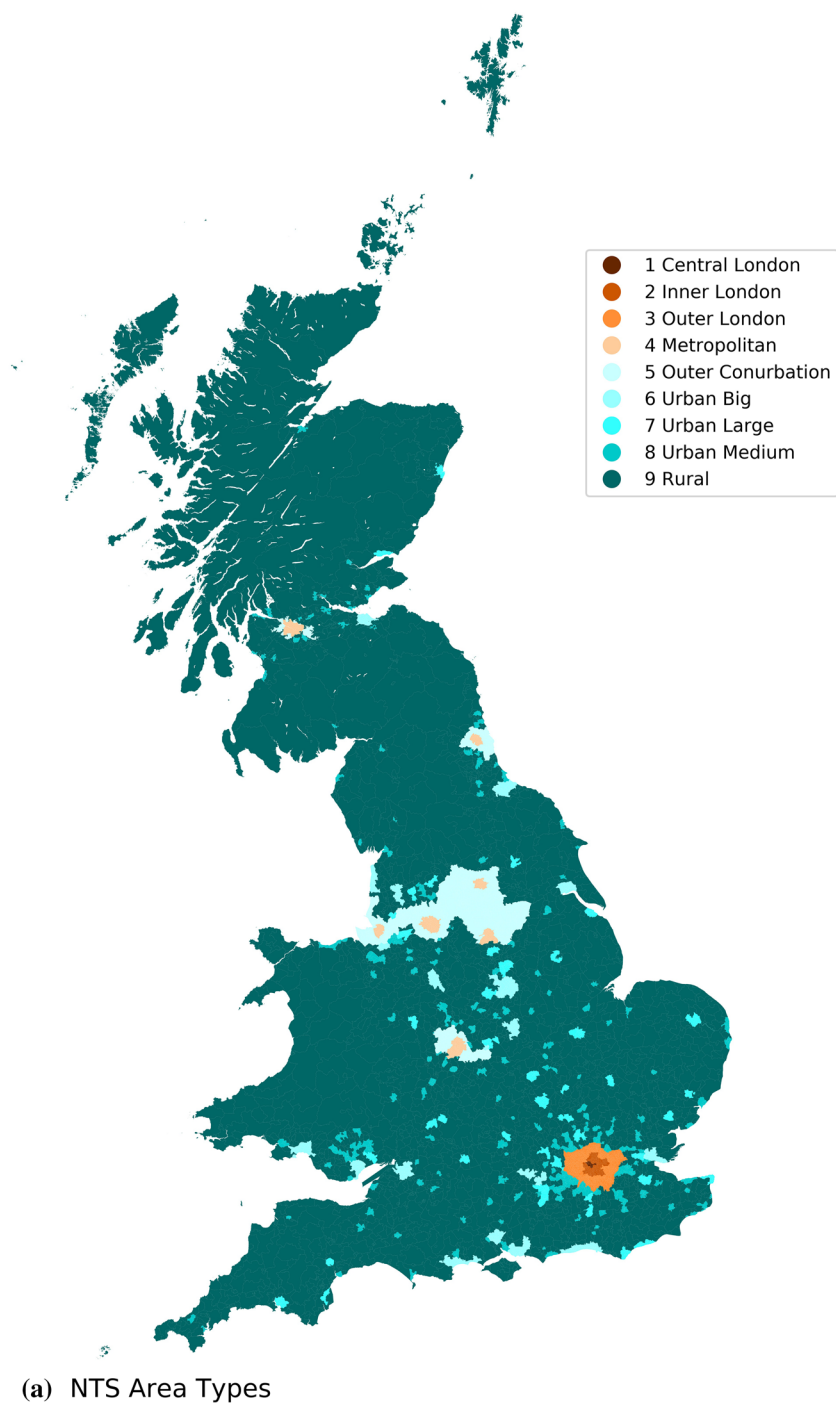
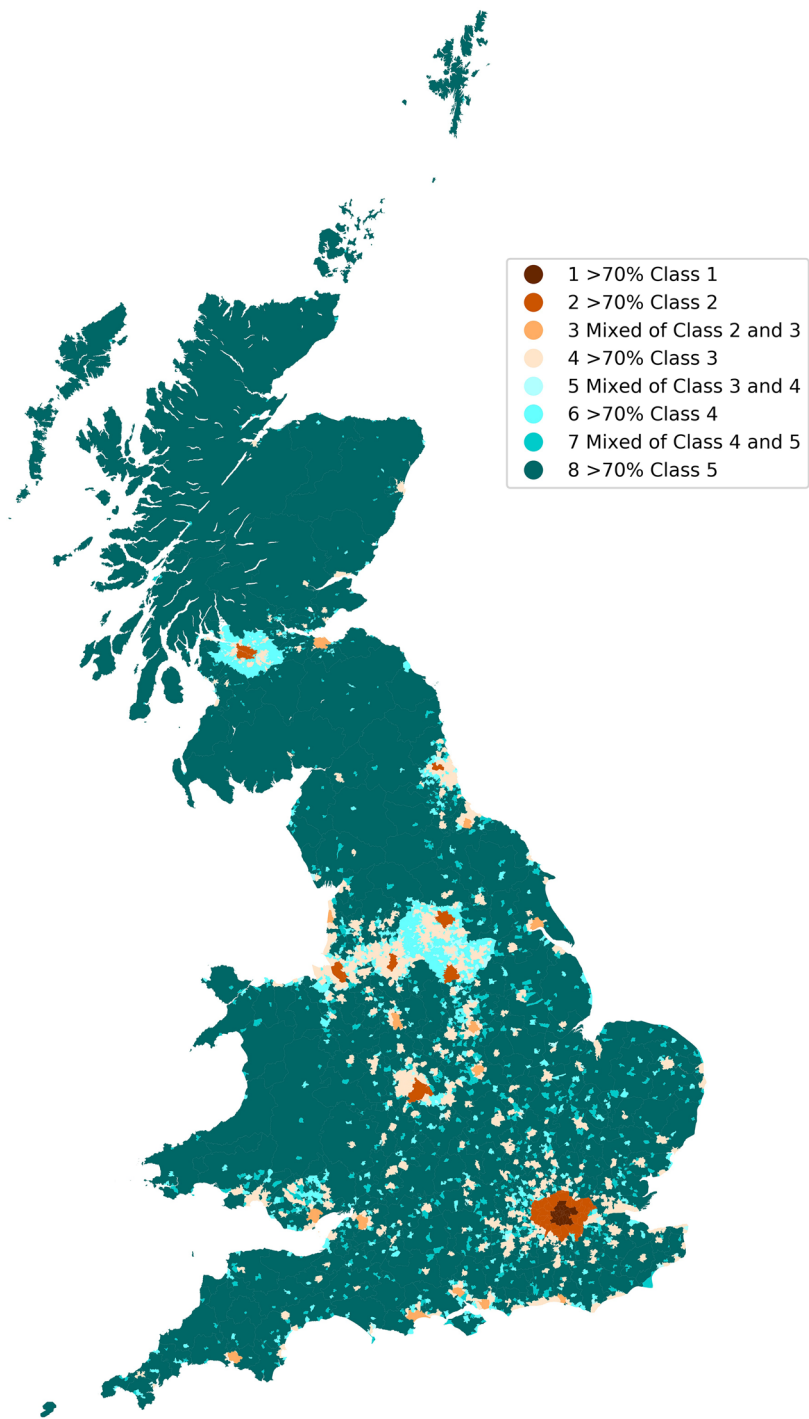


Fig. 3 A map of latent classes in comparison with NTS area type



(b) Spatial distribution of the latent classes

Fig. 3 (continued)

Commuting distance and time elasticity with respect to fuel price

This section applies the developed model to estimate commuting distance and time elasticity with respect to 10% increase in fuel price. Comparing these results against WebTAG guidance serves as a further validation for the estimated model.

Table 11 summarizes the main assumptions made for estimating the fuel price elasticities; part (a) reports the average unleaded petrol cost, 10% of which (i.e. 8.3 pence) used as the assumed increase in fuel price, and the share of travel distance by car which is used to convert the elasticity over all modes to that for cars (i.e. by dividing total elasticity by 70% share of commuting distance by car). Part (b) provides average commuting distance and time for all employed adults; this information is used as the reference values in calculating elasticities.

Table 12 summarize the model response to 10% increase in fuel price; part (a) reports the coefficients for fuel price influences on commuting distance and time extracted from the constrained model for all employed adults. The constrained model is the model for which variations in coefficients across built environment classes are kept constant (i.e. when the average influences across all classes are estimated) These coefficients are then used to calculate changes in travel distance and time as a result of 10% increase in fuel prices. Part (b) to part (d) of Table 12 report these implied changes in terms of absolute value (refer to part b) and percentage change over all modes (part c) and cars (part d).

The results show that the implied car-km elasticities with respect to fuel price are in the same range of those suggested in WebTAG⁶ (refer to part (d) of Table 12). Meeting WebTAG realism test for fuel price is reassuring, suggesting that our model estimation of fuel price is in line with what has been found in existing literature. Additionally, as shown in Sect. 4.1, our models show significant and nonlinear variations in fuel price influences across built environment classes.

Table 11 Information and assumptions for calculating fuel elasticity

Part a) Information extracted from 2015 NTS report ^a	
Average unleaded petrol price (2002 to 2015)	83 pence
Percentage of travel distance for commuting by car	70%
Part b) Information derived from NTS data (2002 to 2015)	
Average commuting distance (weekly)	28.5
Average commuting time (weekly)	79.67

The data represents outbound travel by employed adults in an average 7-day week. It excludes any return trips since the return trips cannot be classified as precisely by travel purposes

^aRefer to page 33 of NTS report (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/551437/national-travel-survey-2015.pdf)

⁶ TAG unit M2: Variable Demand Modelling.

Table 12 Commuting distance and time response to 10% increase in fuel prices

	Average employed person
Part a) Coefficients for fuel price influences on commuting distance and time	
Commuting distance	− 0.07***
Commuting time	− 0.21***
Part b) The effect implied by the coefficients of 10% increase in fuel price over all modes—absolute change	
Commuting distance (miles per week)	− 0.581 ^a
Commuting time (minutes per week)	− 1.743
Part c) The effect of 10% increase in fuel price over all modes—percentage change ^b	
Commuting distance (miles per week)	− 2.04%
Commuting time (minutes per week)	− 2.19%
Part d) The effect of 10% increase in fuel price on car travel—percentage change ^c	
Commuting distance (miles per week)	− 2.9%
Commuting time (minutes per week)	− 3.1%

The data represents outbound travel by employed adults in an average 7-day week. It excludes any return trips since the return trips cannot be classified as precisely by travel purposes

***Significant within 99% CI, **significant within 95% CI, *significant within 90% CI

^aThis is calculated as: fuel cost influence on commuting distance for constrained model (i.e. $-0.007 \times 10\% \times \text{average fuel cost}$ (i.e. 83 pence). The rest are calculated in similar way by using their associated information

^bThis is calculated by dividing absolute change over the associated average distance or time from Table 10

^cThis is calculated by dividing percentage change over all modes by the percentage of commuting distance by car (i.e. 70%). The percentage of car travel distance is assumed to be constant over all income bands

Policy implications

The main purpose of this paper is to develop a new methodology for identifying and mapping spatial variations in travel choices. The empirical test reported above using the UK NTS data shows that this methodology, benefiting from the full range of good-quality travel survey and car fuel price data, reveals new geographical classifications that define distinct land use and built form areas for travel behaviour analyses in more confident ways than in existing literature. For such analyses, leading scholars in the UK have long made it clear their own reservations and criticisms regarding the use of the NTS Area Type (Banister 2002; Headicar 2009; WSP 2005), although the NTS Area Type has so far been the only working option for policy-oriented analyses (Stillwell et al. 2018).⁷

The fact that the analytical findings from this new methodology can pinpoint more precisely the distinctive patterns of travel behaviour means that the SEM–LCA approach can open up new perspectives in policy analyses. The travel behavioural patterns may also be connected with other mapping of human behaviours, such as political voting, political opinions, business clusters, personal and public health, etc. The full implications will take

⁷ Please refer to Paper 4 of https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674568/analysis-from-the-national-travel-survey.pdf.

time to become clearer. Apart from recognising the potential in these areas, it would be difficult for us to pre-judge how far it could go in connecting up the mapping from those diverse disciplines.

Nevertheless, some of the implications for policy are already emerging. These would include:

First, the SEM–LCA methodology provides a new way to identify highly heterogeneous patterns of personal choices across the urban and rural geographies, whilst accounting for self-selection, spatial sorting, endogeneities, and interactions among the explanatory variables. The capability of discerning amongst the diverse segments of travellers is particularly important, especially in the context of continuing growth in the proportion of women, part-time and flexi-time workers, and self-employed workers changes (Palmer 2014; Carl et al. 2014). Those growing segments represent very different travel needs compared with those of the dominant conventional full-time commuters. Insights into the differentiations amongst the traveller segments would inform better transport policies to mitigate social exclusion.

Secondly, the improved geographical precision in mapping the choices and preferences of travel would inform better tailored land use and transport policies at a local level. The model results presented above account for the interdependencies among residents' demographic and socioeconomic profiles, car ownership, travel demand and the built-form. For instance, at the edges of the large metropolitan areas, the approach to sustainable, green and low carbon travel may need to account for the fact that the residents are far more car dependent, and therefore require locally adaptive policies to progress with policy and regulation. The mapping presented above would provide neighbourhood level information for the policy discussions.

Thirdly, the findings above provide concrete geographical classification across Britain and reveal the extent to which the travel patterns vary. This is cogent for defining spatially adapted strategies for planning new transport and land use interventions, particularly in view of the devolved local transport planning and investment responsibilities to municipal and subnational authorities, such as those progressing with the Oxfordshire 2050 Plan (<https://oxfordshireplan.inconsult.uk/consult.ti>), those aiming to implement the Cambridgeshire and Peterborough Independent Economic Review (<https://www.cpier.org.uk/>), and more generally along the Oxford-Milton Keynes—Cambridge Arc, in the West Midlands, Greater Manchester, the Bristol-Bath areas etc.

Fourthly, the findings points to the main challenges in sustainable transport development in the coming decades. They show that it is in LC3 and LC4 (i.e. smaller cities and towns near the main metropolitan areas) where in spite of considerably higher residential population density, those areas have very similar patterns of commuting travel to those living in LC5 (rural areas, mostly remote). Such areas are most frequently found at the edge of the metropolitan areas or in those devolved mayoral authorities which have experienced significant growth, and are aiming to achieve even more jobs and housing growth in the coming decades. Unlike what is commonly believed in the land use planning and urban design circles, in such areas achieving the equivalent residential densities to those observed in LC2 (larger metropolitan built up areas) would be insufficient to alter commuting travel behaviour. What appears to be necessary is to implement systemic changes in land use, built form and transport supply before their commuting patterns converge to those observed in LC1 and LC2.

Conclusions

This paper sets out to investigate geographical patterns of the interdependencies among the main factors that influence car ownership, travel distance and travel time of employed adults in the UK, controlling for a comprehensive range of socioeconomic, demographic and built-form attributes as recorded by the UK National Travel Survey (NTS) and car fuel prices recorded by the Department of Business, Energy & Industrial Strategy. The research has developed a new approach for spatial classification of distinct travel choices. We also set out the main policy implications that have thus far emerged from the analysis.

Methodologically, the research exploits a new combination of structural equation modelling with latent class analysis to analyse a comprehensive set of UK NTS variables. Additionally, it has shown how the methodology allows the incorporation of data from alternative sources (in this case fuel prices as a non-NTS variable). The outcome is the mapping of five specific types of built environment classes that have distinct and non-linear influences upon commuting travel, after rigorous controls of self-selection, spatial sorting and endogeneity effects. We have shown that the new classification have provided new insights in capturing variations in travel choices when compared with the conventional spatial classifiers (e.g. area type which is used as the basis for zonal definition in the current version of the National Transport Model (version 2) and National Trip End Model, as well as a whole range of existing analysis of travel behaviour). Compared with the classification by area type and population density, latent classes offer more distinct identification of travel choices with larger variations of travel across classes.

Building on the technical discussions above, the findings highlights the need for possible new policy interventions in transport and land use planning. The findings show that car ownership and travel choices are highly heterogeneous across settlements. A small proportion (around 8%) of the population have already adopted behaviours typical of Dense Urban areas where car ownership and use have become secondary, but the majority behave in a similar way to those in Rural areas which are highly car dependent. The most interesting cases are, however, the smaller cities and towns which are found in the peripheries of metropolitan conurbations and urban settlements down to 25,000 people. The mapping of the area classes show that these areas include many of the innovative, high job growth areas like the M4 corridor outside London, Oxford, Milton Keynes, Bedford, Northampton, Peterborough and the M11 Corridor up to Cambridge, where major transport and land use initiatives are being launched to support sustainable development. We expect that it is in these high growth areas that the SEM–LCA will pick up car ownership and travel behaviour change in the future through NTS and other sources.

Although this research has postulated the first attempt in mapping and incorporating the geographical dimension in analyzing the heterogenous influences on travel behaviour, we are only at the start of exploiting the potential of this new method. First, the method can be further enhanced through the inclusion of more non-NTS and geographically specific variables in future work. For instance, the latent classification can create a basis for data fusion through mapping the data available at more aggregate spatial level. Second, the spatial classification of the human behaviors and the developed SEM–LCA model can open up the perspective for new policy analyses within and outside the transport sector (e.g. environment, public health, political opinion, social exclusion, etc.).

Acknowledgements This paper is the extended and improved version of that presented at 2018 Transportation Research Board annual meeting. On behalf of both authors, the corresponding author states that there is no conflict of interest. The views and opinions expressed in this article are those of the authors and do

not necessarily reflect the official policy or position of the UK Department for Transport who have sponsored the work, and any other research support. The authors wish to acknowledge the support from the UK Department for Transport's Transport-Technology Research Innovation Grant (T-TRIG). In addition, Ying Jin would like to acknowledge the funding support from the UK EPSRC Centre for Smart Infrastructure and Construction Phase 2 (EP/N021614/1) and a special fund from Key Laboratory of Eco Planning and Green Building, Ministry of Education (Tsinghua University), China.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Banister, D.: Transport Policy and the Environment. Routledge, London (2002)
- Beckman, J.D., Goulias, K.G.: Immigration, residential location, car ownership, and commuting behavior: a multivariate latent class analysis from California. *Transportation* **35**, 655–671 (2008)
- Bohte, W., Maat, K., Van Wee, B.: Measuring attitudes in research on residential self-selection and travel behaviour: a review of theories and empirical research. *Transp. Rev.* **29**, 325–357 (2009)
- Cao, X., Mokhtarian, P.L., Handy, S.L.: Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transp. Rev.* **29**, 359–395 (2009)
- Carl, E., Heald, K., Hood, A.: The Changing Face of Retirement: Future Patterns of Work, Health, Care and Income Among the Older Population. Institute for Fiscal Studies, London (2014)
- De Haas, M.C., Scheepers, C.E., Harms, L.W.J., Kroesen, M.: Travel pattern transitions: applying latent transition analysis within the mobility biographies framework. *Transp. Res. A Policy Pract.* **107**, 140–151 (2018)
- De Ona, J., López, G., Mujalli, R., Calvo, F.J.: Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. *Accid. Anal. Prev.* **51**, 1–10 (2013)
- Delbosc, A., Naznin, F.: Future life course and mobility: a latent class analysis of young adults in Victoria, Australia. *Transp. Policy* **77**, 104–116 (2019)
- Ding, C., Cao, X., Wang, Y.: Synergistic effects of the built environment and commuting programs on commute mode choice. *Transp. Res. A Policy Pract.* **118**, 104–118 (2018)
- Gao, S., Mokhtarian, P.L., Johnston, R.A.: Exploring the connections among job accessibility, employment, income, and auto ownership using structural equation modeling. *Ann. Reg. Sci.* **42**, 341–356 (2008)
- Handy, S., Cao, X.Y., Mokhtarian, P.: Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transp. Res. D Transp. Environ.* **10**, 427–444 (2005)
- Headicar, P.: Transport Policy and Planning in Great Britain. Routledge, London (2009)
- Headicar, P.: Variations in Travel Between Different Locations. UK Department for Transport, pp 12–19 (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674568/analysis-from-the-national-travel-survey.pdf. Accessed 1 June 2019
- Jahanshahi, K., Jin, Y.: The built environment typologies in the UK and their influences on travel behaviour: new evidence through latent categorisation in structural equation modelling. *Transp. Plan. Technol.* **39**, 59–77 (2015)
- Jahanshahi, K., Jin, Y.: Trendbreaking influences of built form on travel in UK cities: evidence from new quantifications of within- and between-built-form variations. *Transp. Res. Rec. J. Transp. Res. Board* **2564**(1), 31–40 (2016)
- Jahanshahi, K., Jin, Y., Williams, I.: Direct and indirect influences on employed adults' travel in the UK: new insights from the National Travel Survey data 2002–2010. *Transp. Res. A Policy Pract.* **80**, 288–306 (2015)

- Jahanshahi, K., Williams, I., Jin, Y.: Understanding rail commuting market segments in Britain through latent class analysis. In: Transportation Research Board 98th Annual Meeting. Washington DC, United States (2019)
- Liao, F.H., Farber, S., Ewing, R.: Compact development and preference heterogeneity in residential location choice behaviour: a latent class analysis. *Urban Stud.* **52**, 314–337 (2015)
- Maldonado-Hinarejos, R., Sivakumar, A., Polak, J.W.: Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. *Transportation* **41**, 1287–1304 (2014)
- Mokhtarian, P.L., Cao, X.: Examining the impacts of residential self-selection on travel behavior: a focus on methodologies. *Transp. Res. B Methodol.* **42**, 204–228 (2008)
- Morris, S., Humphrey, A., Pickering, K., Tipping, S., Templeton, I., Hurn, J.: National Travel Survey 2013 Technical Report. UK Department for Transport (2014)
- Palmer, N.: Trends in self employment. ONS Economic Forum (2014)
- Robert, C., Murakami, J.: Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environ. Plan. A* **42**, 400–418 (2010)
- Silva, J.D.A.E., Morency, C., Gouliasc, K.G.: Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transp. Res. A Policy Pract.* **46**, 1252–1264 (2012)
- Stillwell, D., Cummings, J., Slocombe, M.: Analyses from the National Travel Survey. In: DfT, UK. (ed.) (2018)
- Sun, Y., Waygood, E., Fukui, K., Kitamura, R.: Built environment or household life-cycle stages—Which explains sustainable travel more? *Transp. Res. Rec. J. Transp. Res. Board* **2135**, 123–129 (2009)
- Sun, Y., Waygood, E., Huang, Z.: Do automobility cohorts exist in urban travel? *Transp. Res. Rec. J. Transp. Res. Board* **2323**, 18–24 (2012)
- Van Acker, V., Witlox, F., Van Wee, B.: The effects of the land use system on travel behavior: a structural equation modeling approach. *Transp. Plan. Technol.* **30**, 331–353 (2007)
- Wang, J., Wang, X.: *Structural Equation Modeling: Applications Using Mplus*. Wiley, Hoboken (2012)
- Wang, T., Chen, C.: Attitudes, mode switching behavior, and the built environment: a longitudinal study in the Puget Sound Region. *Transp. Res. A Policy Pract.* **46**, 1594–1607 (2012)
- WSP: Impacts of Land Use Planning Policy on Transport Demand and Congestion. DfT, Cambridge (2005)
- Zegras, C., Lee, J.S., Ben-Joseph, E.: By community or design? Age-restricted neighbourhoods, physical design and baby boomers' local travel behaviour in Suburban Boston, US. *Urban Stud.* **49**, 2169–2198 (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Kaveh Jahanshahi is currently working to a combined academic-civil service career, as Senior Lecturer in data science at the UK Office of National Statistics and Research Associate in urban analysis and transport modelling at the Department of Architecture, University of Cambridge, with both research and teaching responsibilities. In recent years, his research has focused mainly on understanding urban issues through digital urban modelling and use of advanced statistical and machine learning models in the UK and across different geographic areas of the world. Prior to taking on lecturing and academic career, Kaveh was working as Transport Modelling Team Lead and senior Transport Modeller at the Government and for different global consultants with technical leading responsibilities. Managing the Department for Transport's "National Transport Model", and APRIL, which models the impact of transport policies in London, are two examples of the strategic modelling developments which are lead by Kaveh.

Dr. Ying Jin lectures on city planning, urban design, and urban modelling. He is particularly interested in understanding how technology, policy and human behaviour affect the development of cities and their infrastructure, and in using this knowledge to create new design solutions. His main research interests are computer models of cities, and urban history. He has extensive industry experience and directed multi-disciplinary teams in building and using computer models as experimental platforms to appraise policy scenarios that involve investment, regulation, pricing and promotional campaigns. Key projects include strategic planning of London and surrounding regions, sub-regional and local planning in the English Midlands, transport and energy scenarios for the European Union, long term city region and transport plans in China and in South America, mapping urban poverty in emerging economies, and assessing development and transport options for Cambridge and surrounding regions.